Good Cop, Bad Cop

An Evaluation of Use-Of-Force Complaints Made Against the Chicago Police

Mihir Bhaskar, Abhilash Biswas, and Jameson Carter

The Problem

The Chicago PD have a *forceful* reputation.

Policymakers must:

- Target efforts towards areas where use-of-force complaints proliferate
- Evaluate underlying reasons for differing complaint outcomes - why are some complaints upheld and others struck down?

Chicago PD, using force at a peaceful protest, 2020

798

Census Tracts

With extensive demographic information collected by the American Community Survey

647

Use of Force Complaints

Made by residents against the Chicago Police, from 2015-2018

8

Major Roads

And other transit features such as railways that often delineate neighborhoods

23k

Weapons Violations

Charged against residents of Chicago, alongside other crime data

5

Rivers and Canals

And other hydrological features such as Lake Michigan

27k

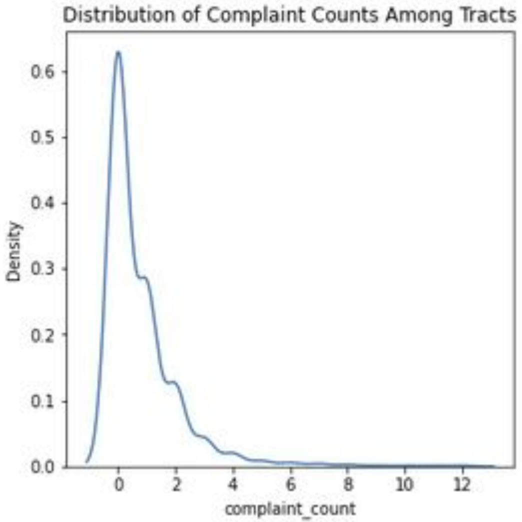
Investigatory Stops

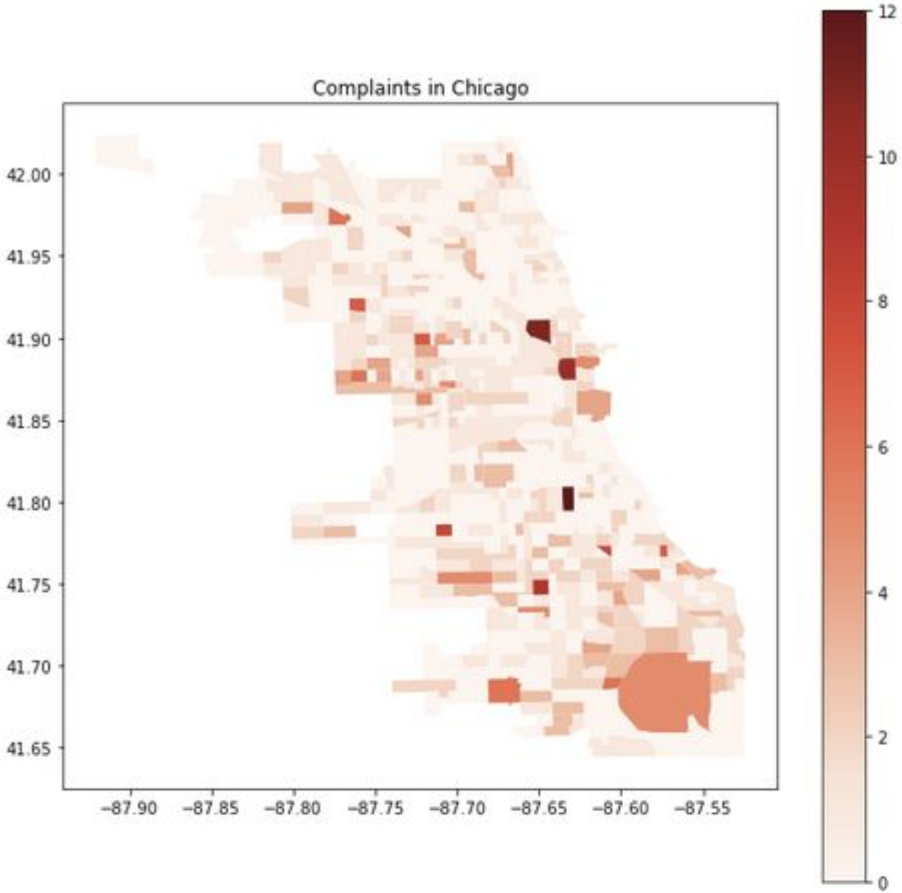
Conducted by the Chicago Police, of residents from 2016-2019

Exploratory Data Analysis

Distribution of Complaints

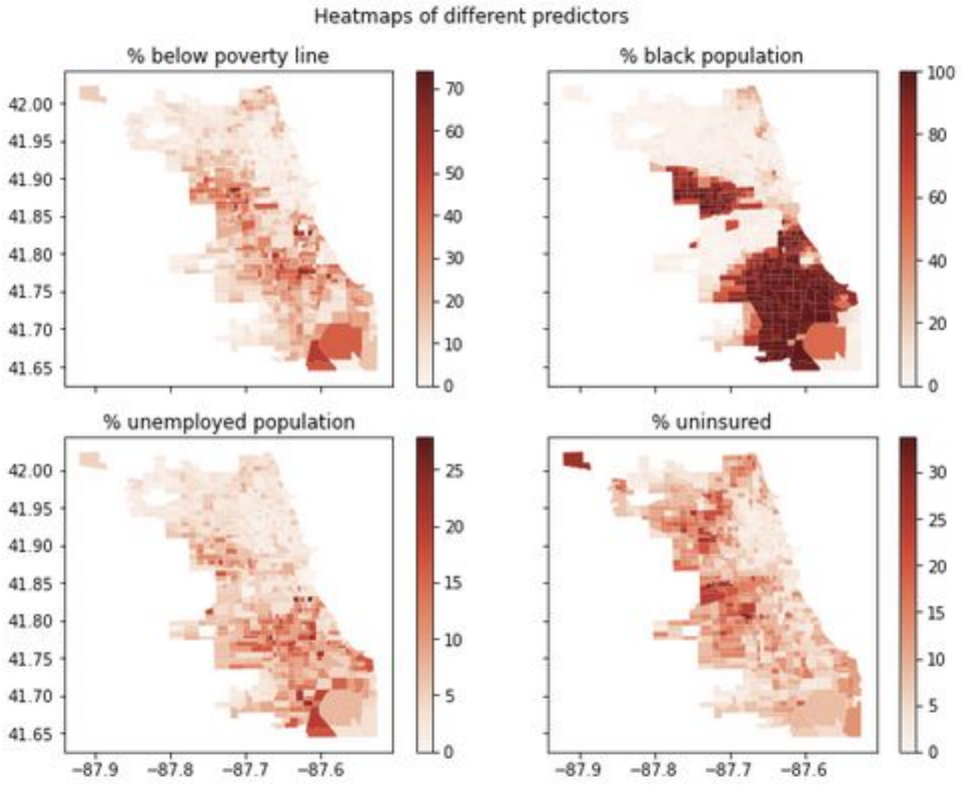
Highly skewed! Some selected areas with high counts...





Predictors geographically vary

The South Side has a higher Black population, generally more poverty and unemployment.

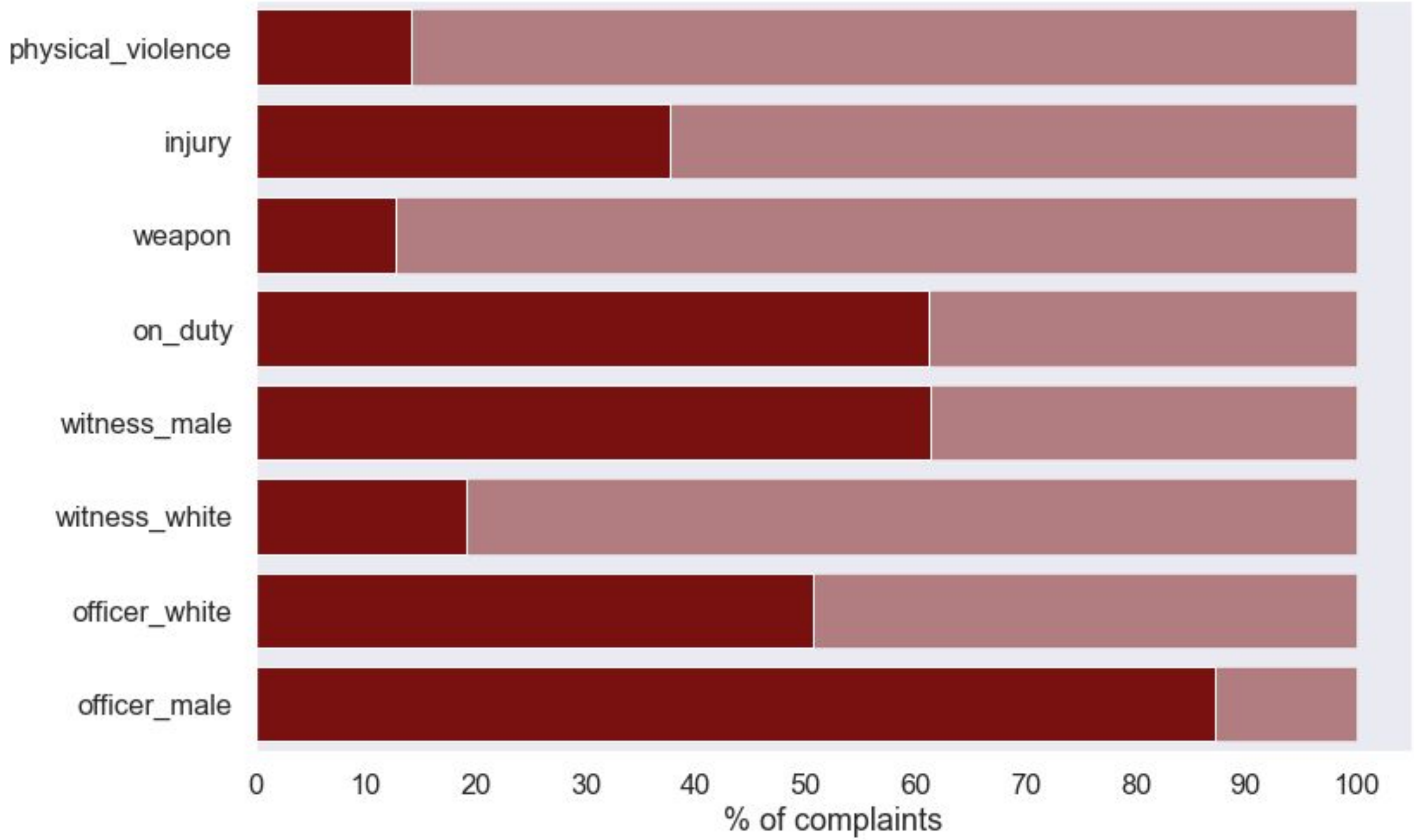


There is sufficient variation in complaint characteristics

About the **incident**

About the **victim**

About the **officer**



The three research questions:

1. What dictates complaints in Census Tracts? Can we predict or classify their occurence?

2. How do these communities differ from one another? Can we cluster tracts effectively, and do complaints vary across clusters?

3. What dictates the outcome of a complaint? Can we predict or classify these outcomes?

Research Question 1: Predicting and Classifying Number of Complaints in Census Tracts

What is our outcome variable?

Prediction Task —>

Count of use-of-force complaint at a Census tract level

	geo_id	complaint_count
793	1400000US17031843500	1.0
794	1400000US17031843600	1.0
795	1400000US17031843700	3.0
796	1400000US17031843800	1.0
797	1400000US17031843900	2.0

Classification Task —>

A binary outcome of any complaint (1 if there has been at least 1 complaint in a given Census tract, and 0 otherwise)

	any_complaints
0	454
1	344

Predicting exact count of complaints is challenging

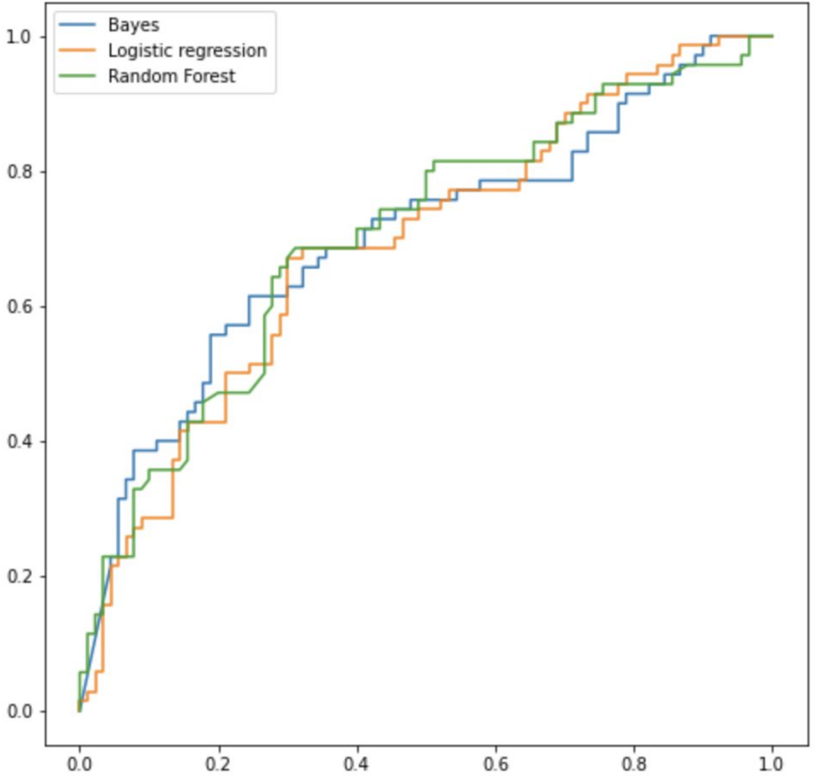
	Model	Predictors	R square on validation set
0	Linear OLS	All	18.50%
1	Linear OLS	Best 40	~0%
2	Ridge	All	~12%
3	Lasso	All	~18.5%
4	Random Forest Regressor	All	~19%

- *Best 40 predictors chosen by forward selection*
- *Ridge shrinkage method was applied with a best alpha of 10000*
- *Lasso shrinkage method was applied with a best alpha of 10.23*

Classification models perform better than prediction models

Model	Accuracy	Recall (0)	Recall (1)	AUC
Random Forest	70%	73.30%	66%	0.7
Naïve Bayes	66%	86%	41%	0.70
Logistic Regression	66%	79%	49%	0.69
KNN	59%	78%	36%	

ROC

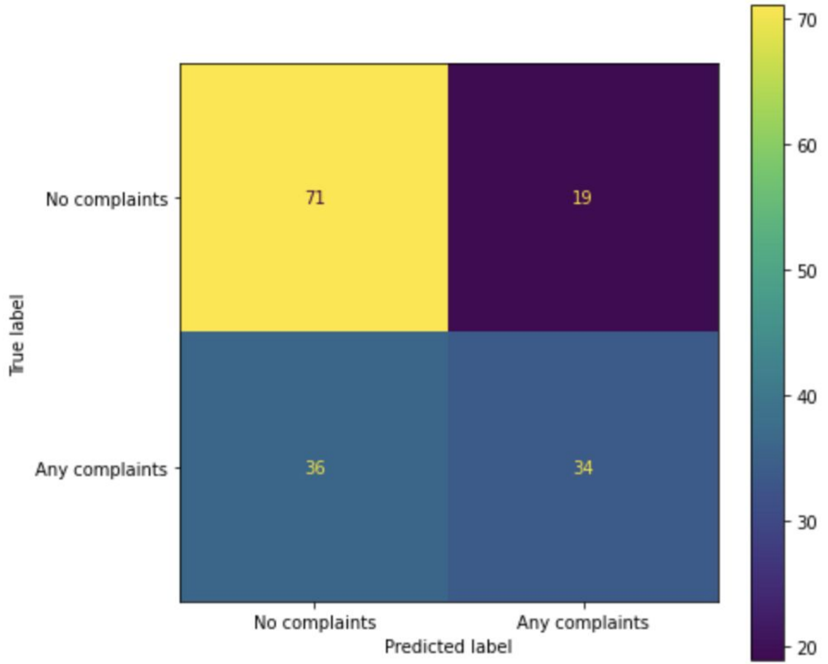


Tuning the best classification model: Random Forest

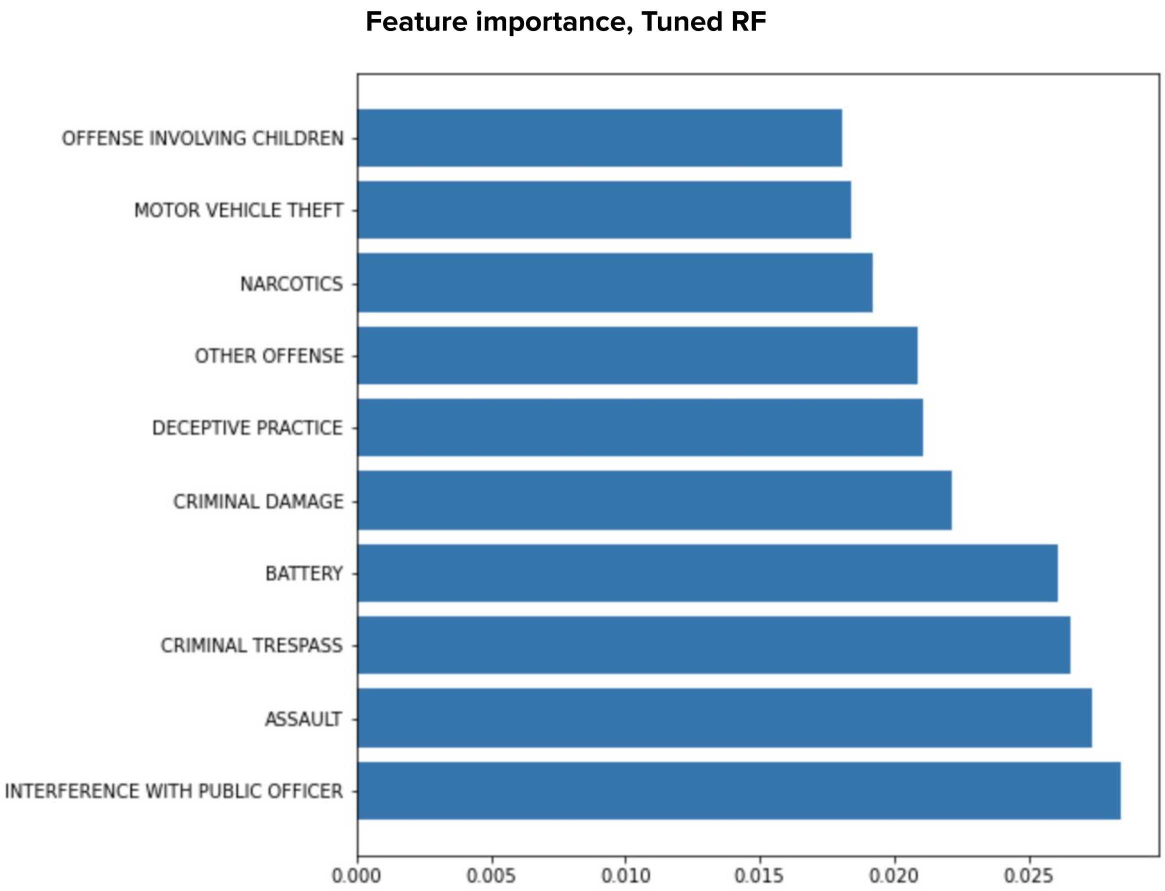
Tuned parameters: max_depth=30, min_samples_split=6, n_estimators=400

Class	Precision	Recall
No complaints	73%	73%
Any complaints	66%	66%

Overall accuracy: 70%



Most important variable - Interference with public officer



However, even the most important variables have low correlation with outcome variable

Variable	Correlation with any complaints
Interference with public officer	0.34
Assaults	0.38
Criminal trespass	0.28
Battery crime violations	0.39

Key takeaways

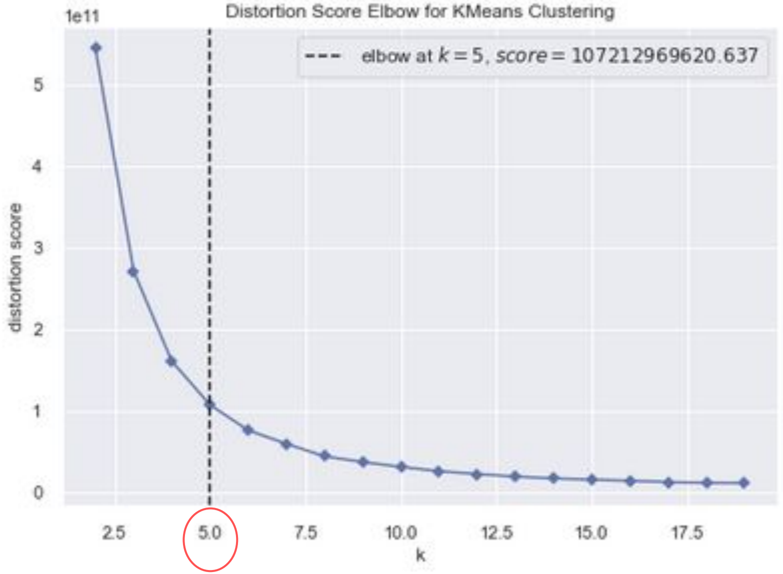
- Predicting exact complaint count is much more difficult than classifying whether a tract has any complaint or not

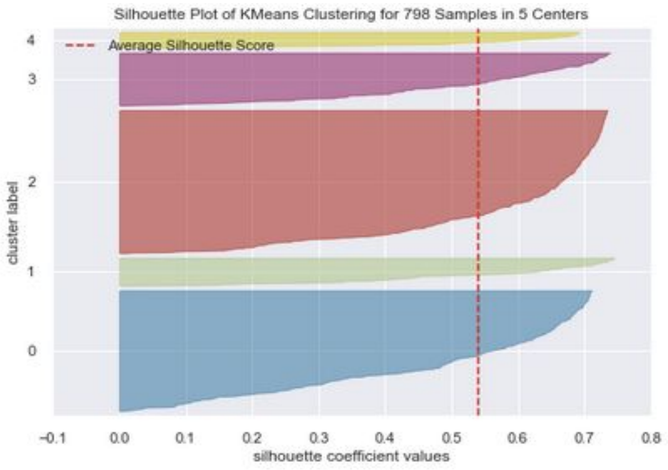
- Classification performance is worse on tracts where there has been atleast 1 complaint (any_complaint = 1)

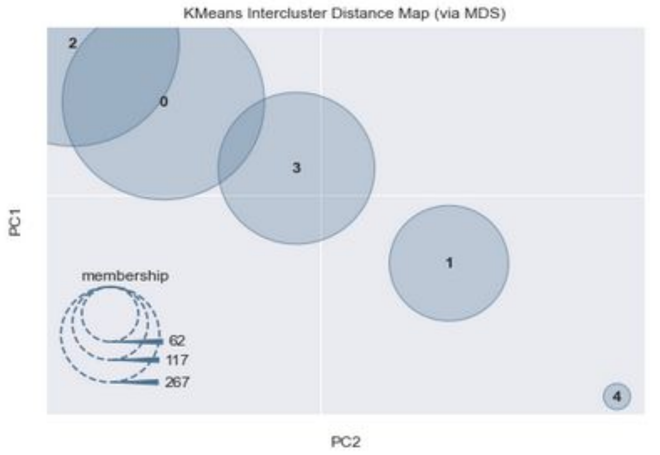
- Even the most important variables have a low correlation with outcome variable, which is a key limitation of these models.

Research Question 2: Clustering Census Tracts' Characteristics

Can we observe K-means clusters in our data?







Initially... no

The curse of dimensionality? PCA to the rescue

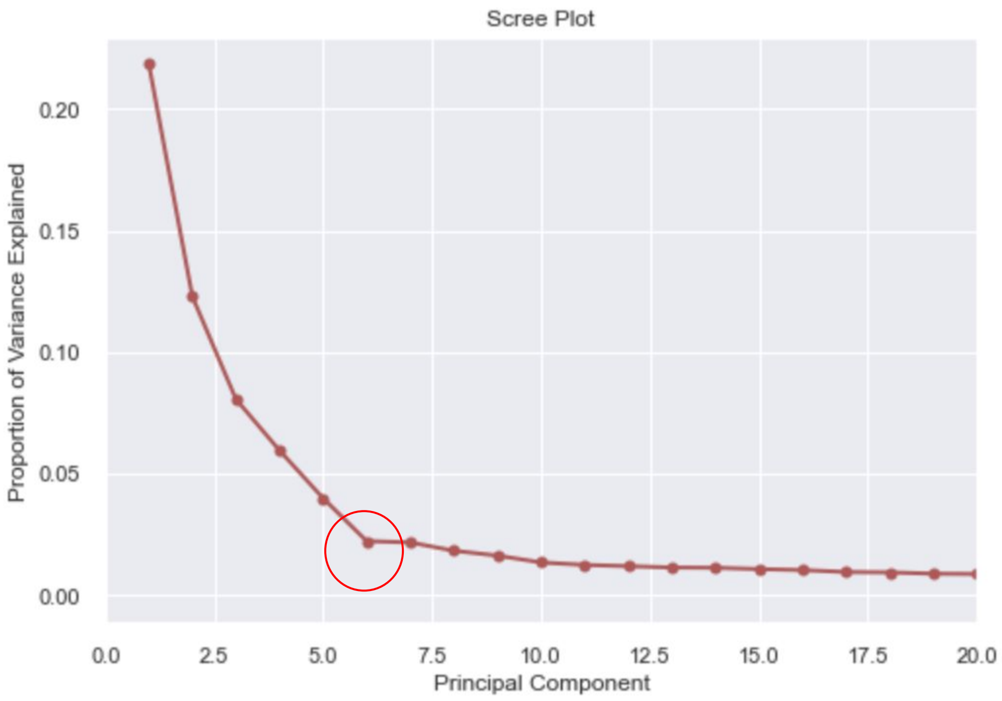
Allows us to reduce the dimensionality of the data. PCA is particularly useful in situations like ours, where there are many highly correlated variables in the dataset..

Col1	Col2	Correlation Coefficient
ASSAULT	BATTERY	0.963624
DECEPTIVE PRACTICE	THEFT	0.953793
ASSAULT	OTHER OFFENSE	0.940364
DP02_0015PE	DP03_0066PE	0.940157
BATTERY	OTHER OFFENSE	0.936715
DP03_0004PE	DP03_0013PE	0.925954
DP05_0037PE	DP05_0038PE	-0.923312
DP02_0093PE	DP02_0114PE	0.917271
BATTERY	CRIMINAL DAMAGE	0.914779
CRIMINAL DAMAGE	OTHER OFFENSE	0.913783
ASSAULT	CRIMINAL DAMAGE	0.911788
DP03_0074PE	DP03_0098PE	0.896063
DP02_0014PE	DP05_0019PE	0.886578
INTERFERENCE WITH PUBLIC OFFICER	WEAPONS VIOLATION	0.885598

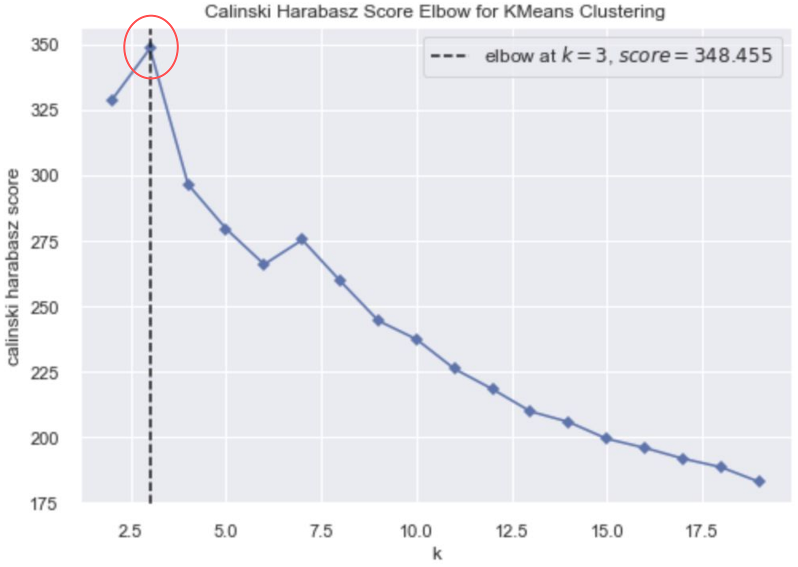
And the list goes on!

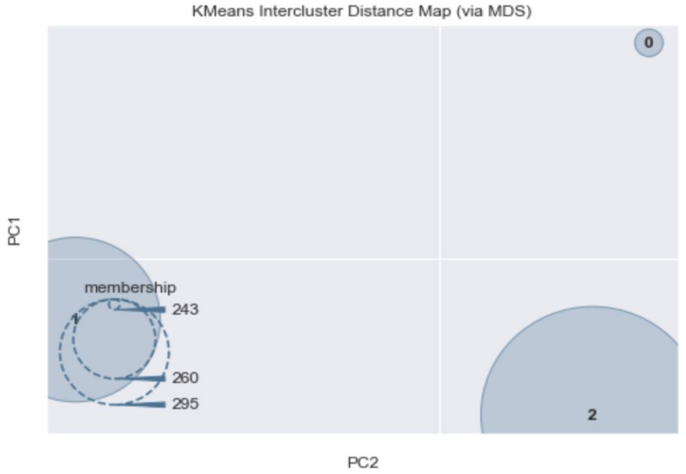
But how many principal components do we use?

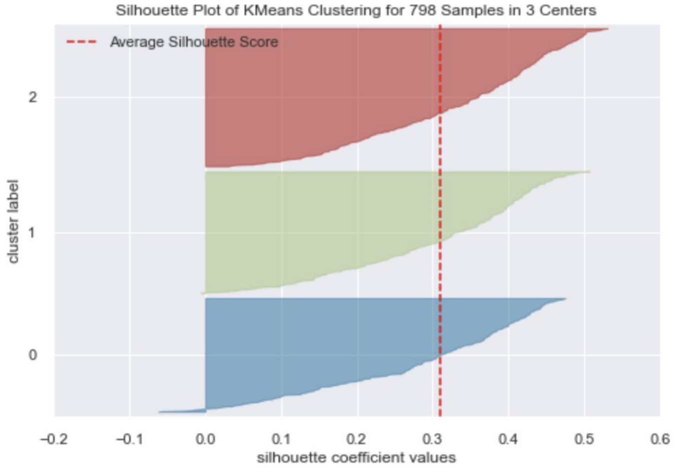
It seems that 6 principal components explain most of the variation in the dataset.



Now cluster using principal components as inputs! 3 looks promising...







Final model alert!

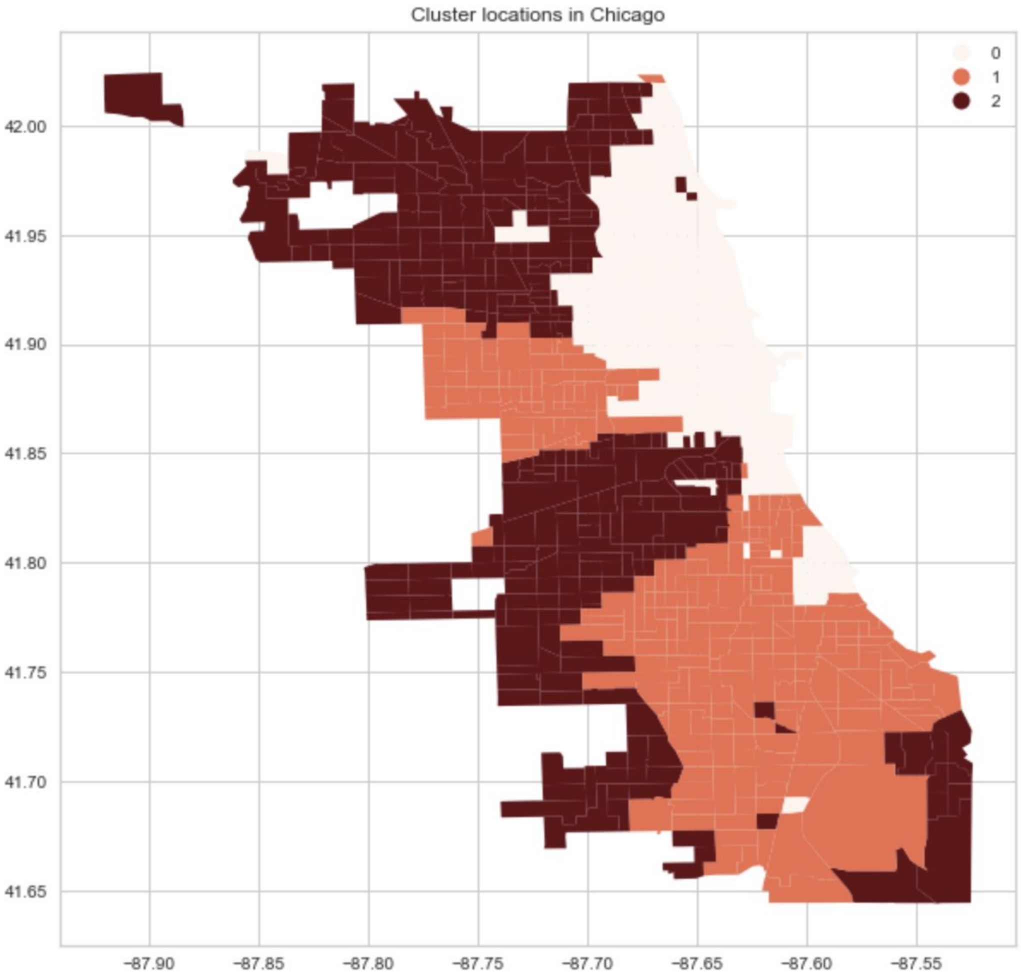
K-means clustering over six principal components, where K = 3!

Component	Most Influential Variable
PC1	% with public health insurance
PC2	Mean household size
PC3	Total population
PC4	# Older folks with a disability
PC5	% of population born abroad
PC6	# Enrollees in school

Finding 1.

One cluster is drawn sharply on racial lines, and median complaints there are higher:

Cluster	Median Black %	Median Complaints
0	5.5%	0
1	92.7%	1
2	3.2%	0

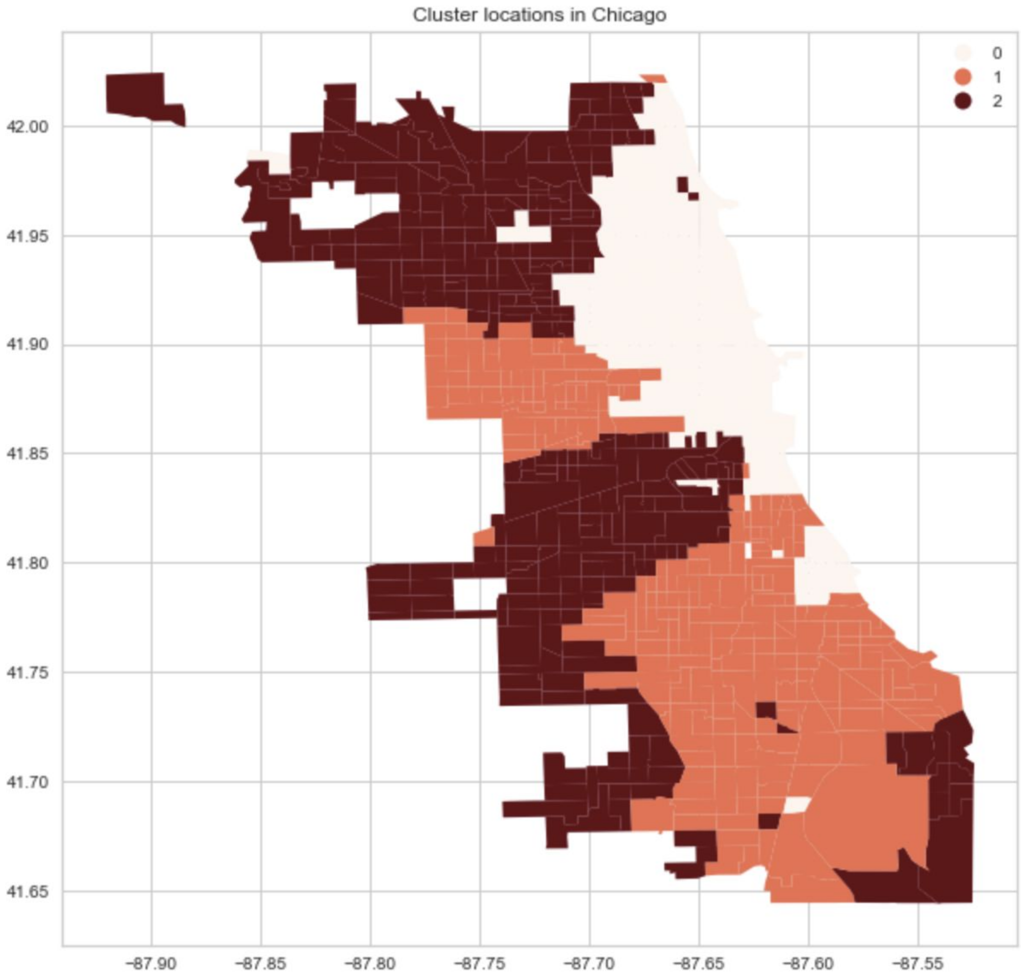


Component	Most Influential Variable	Cluster: 0	1	2
PC1	% with public health insurance	17.3%	58.1%	37.0%
PC2	Mean household size	2.1	2.6	3
PC3	Total population	2,936	2,418	4,019
PC4	# Older folks with a disability	231	324	465
PC5	% of population born abroad	14.9%	2.6%	33.0%
PC6	# Enrollees in school	564	351	695

Finding 2.

The cluster with the highest use of force complaints at the median:

- Has a higher median black population and utilization of public insurance.
- Has a smaller median population and percentage of foreign-born residents, and fewer school enrollees.



Research Question 3: Classifying Complaint Findings

What are we trying to classify?

A complaint can lead to many possible outcomes after investigation

‘Sustained’ vs. ‘Not Sustained’

- Clear interpretation of target
- Balanced classes

Finding	Number of complaint-officer pairs
No Affidavit	398
Unfounded	251
Unknown	187
Not Sustained	162
Sustained	112
Exonerated	70

The curse of dimensionality... revisited using VIF

Instead of using PCA, reduce the number of variables by using the **Variance Inflation Factor**

Keep only variables where VIF <= 5 as a 'subset'

Col1	Col2	Correlation Coefficient
ASSAULT	BATTERY	0.963624
DECEPTIVE PRACTICE	THEFT	0.953793
ASSAULT	OTHER OFFENSE	0.940364
DP02_0015PE	DP03_0066PE	0.940157
BATTERY	OTHER OFFENSE	0.936715
DP03_0004PE	DP03_0013PE	0.925954
DP05_0037PE	DP05_0038PE	-0.923312
DP02_0093PE	DP02_0114PE	0.917271
BATTERY	CRIMINAL DAMAGE	0.914779
CRIMINAL DAMAGE	OTHER OFFENSE	0.913783
ASSAULT	CRIMINAL DAMAGE	0.911788
DP03_0074PE	DP03_0098PE	0.896063
DP02_0014PE	DP05_0019PE	0.886578
INTERFERENCE WITH PUBLIC OFFICER	WEAPONS VIOLATION	0.885598

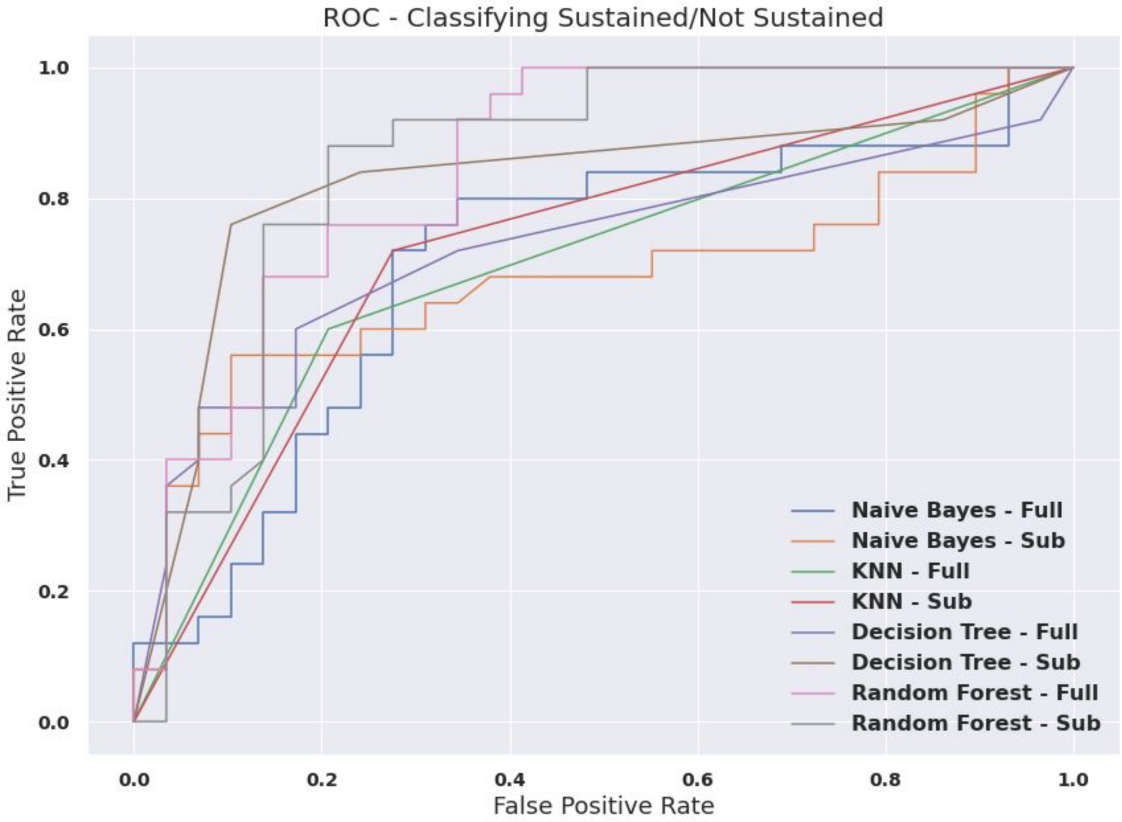
Choosing the best model

Feature sets:

- Full: all features
- Sub: features with VIF <= 5

Best models, by AUC:

1. **Random Forest Sub (0.86)**
2. Decision Tree Sub (0.82)
3. KNN Sub (0.72)
4. Naive Bayes Sub (0.67)

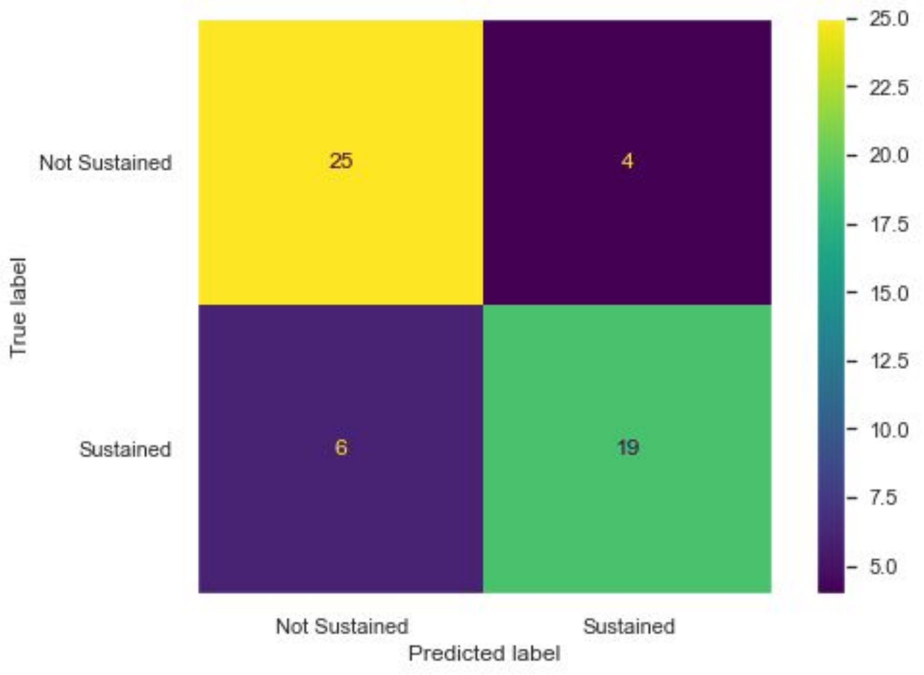


GridSearch tuning and final validation

Tuned parameters: max_depth=None, min_samples_split=2, n_estimators=100

Class	Precision	Recall
Not Sustained	81%	86%
Sustained	83%	76%

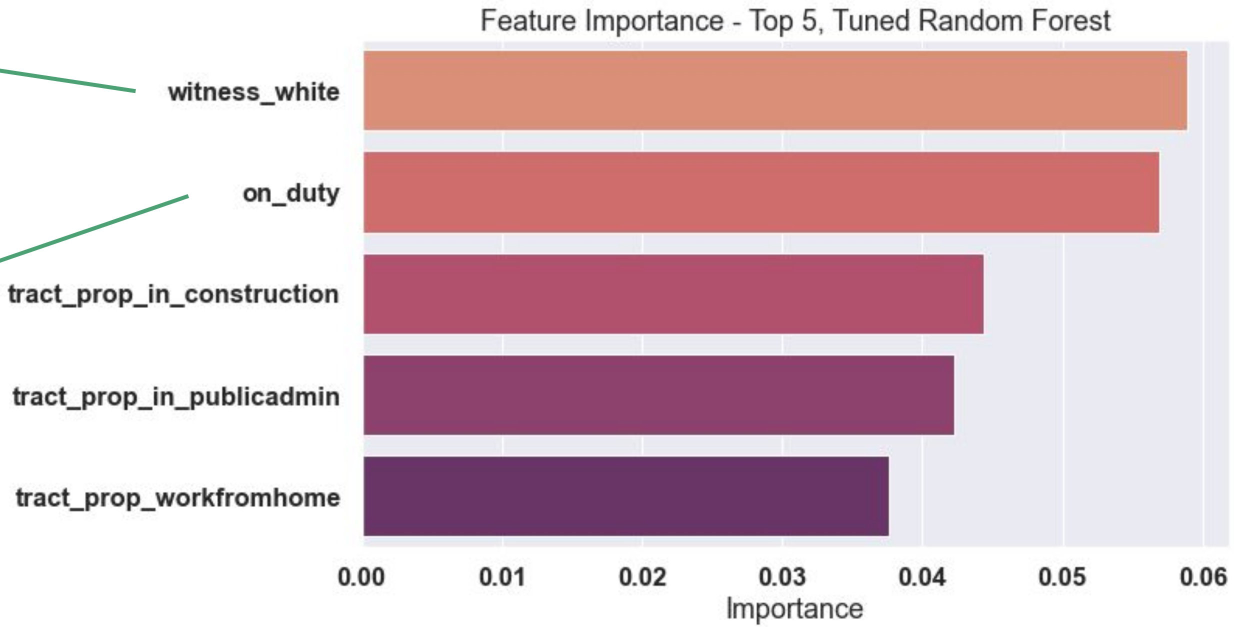
Overall accuracy: 82%



Important features

71% complaints sustained when witness was white, 29% when not

60% complaints sustained when incident happened off-duty, 22% when on-duty



Conclusions and Discussion

Revisiting the three research questions:

1. What dictates the occurrence of complaints in Census Tracts?

   Classified with 70% accuracy. Number of reported crimes matter.

2. How do these communities differ from one another?

   One particular cluster in the data experiences higher complaints at the median; that cluster has a higher black population, poverty rates, and public insurance rates.

3. What dictates the outcome of a sustained complaint?

   Classified with 82% accuracy. Witness race and on/off-duty matter!

Limitations

- We do not have police officer or police precinct data which are likely important predictors of use of force.
- Use of force complaints is a proxy of actual police violence and can be biased.
- It is entirely possible that these incidents are under-reported!

Future Work

If given more time, we might...

- Add features
    - Additional features on police officer and police precinct characteristics
    - Richer geographic features, tuning distances from rivers, lakes, roads *et cetera*
- Model with greater depth
    - Affidavit vs. no affidavit -> understand what determines whether people take complaints forward or just drop them
    - Modelling on other metrics of police violence (civilian deaths, civilian injuries, other categories of complaints)
    - Examine tracts which do not fall neatly into the clusters described in research question 2.
    - Try other initialization methods for K-means clustering model.

Image Sources

- https://www.shutterstock.com/image-vector/nightstick-icon-vector-illustration-design-police-777974059
- https://www.vectorstock.com/royalty-free-vector/police-car-chasing-another-car-icon-vector-31138816
- https://www.shutterstock.com/image-vector/gun-icon-logo-isolated-sign-symbol-1723381117
- https://www.wcbu.org/state-news/2016-09-09/illinois-asks-federal-government-for-medicaid-flexibility
- https://www.vectorstock.com/royalty-free-vector/population-people-icon-design-vector-22023740
- https://www.shutterstock.com/image-vector/old-man-silhouette-1216992988
- https://designbundles.net/hardqor4ik/732180-book-icon-symbol-isolated-on-white-background

References

- https://geopandas.org/en/stable/gallery/create_geopandas_from_pandas.html
- https://stackoverflow.com/questions/38372188/how-to-add-k-means-predicted-clusters-in-a-column-to-a-dataframe-in-python
- https://stackoverflow.com/questions/41128456/pandas-filter-across-all-columns
- https://stackoverflow.com/questions/33575587/pandas-dataframe-how-to-apply-describe-to-each-group-and-add-to-new-columns
- In class labs and homeworks on prediction and classification
- https://scikit-learn.org/stable/modules/generated/sklearn.metrics.auc.html
- https://stackoverflow.com/questions/32860849/classification-pca-and-logistic-regression-using-sklearn
- My in-class exercise on clustering: Clustering-Mini2_JamesonCarter
- https://dzone.com/articles/kmeans-silhouette-score-explained-with-python-exam
- https://365datascience.com/tutorials/python-tutorials/pca-k-means/
- https://dl.acm.org/doi/10.1145/1015330.1015408
- https://stackoverflow.com/questions/22984335/recovering-features-names-of-explained-variance-ratio-in-pca-with-sklearn
- https://medium.com/swlh/k-means-clustering-on-high-dimensional-data-d2151e1a4240